JOURNAL OF CLINICAL MICROBIOLOGY, Dec. 2007, p. 3891–3902 Vol. 45, No. 12
0095-1137/07/$08.00+0 doi:10.1128/JCM.01394-07

Discovery of a Novel *Mycobacterium tuberculosis* Lineage That Is a Major Cause of Tuberculosis in Rio de Janeiro, Brazil▽†

Luiz Claudio Oliveira Lazzarini,[1,2]‡ Richard C. Huard,[1]‡§ Neio L. Boechat,[2] Harrison M. Gomes,[3] Maranibia C. Oelemann,[3] Natalia Kurepina,[4] Elena Shashkina,[4] Fernanda C. Q. Mello,[2] Andrea L. Gibson,[1] Milena J. Virginio,[3] Ana Grazia Marsico,[2] W. Ray Butler,[5] Barry N. Kreiswirth,[4] Philip N. Suffys,[3] Jose Roberto Lapa e Silva,[1,2] and John L. Ho[1]*

Department of Medicine, Division of International Medicine and Infectious Diseases, Weill Medical College of Cornell University, New York, New York 10021[1]; Institute of Thoracic Diseases, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, 21941-590[2]; Laboratory of Molecular Biology Applied to Mycobacteria, Oswaldo Cruz Institute, Fiocruz, Rio de Janeiro, Brazil, 21040-900[3]; Public Health Research Institute, Newark, New Jersey 07103[4]; and Centers for Disease Control and Prevention, Atlanta, Georgia 30333[5]

Received 11 July 2007/Returned for modification 1 September 2007/Accepted 17 September 2007

The current study evaluated *Mycobacterium tuberculosis* isolates from Rio de Janeiro, Brazil, for genomic deletions. One locus in our panel of PCR targets failed to amplify in ~30% of strains. A single novel long sequence polymorphism (>26.3 kb) was characterized and designated RD^{Rio}. Homologous recombination between two similar protein-coding genes is proposed as the mechanism for deleting or modifying 10 genes, including two potentially immunogenic PPE proteins. The flanking regions of the RD^{Rio} locus were identical in all strains bearing the deletion. Genetic testing by principal genetic group, spoligotyping, variable-number tandem repeats of mycobacterial interspersed repetitive units (MIRU-VNTR), and IS*6110*-based restriction fragment length polymorphism analysis cumulatively support the idea that RD^{Rio} strains are derived from a common ancestor belonging solely to the Latin American-Mediterranean spoligotype family. The RD^{Rio} lineage is therefore the predominant clade causing tuberculosis (TB) in Rio de Janeiro and, as indicated by genotypic clustering in MIRU-VNTR analysis, the most significant source of recent transmission. Limited retrospective reviews of bacteriological and patient records showed a lack of association with multidrug resistance or specific risk factors for TB. However, trends in the data did suggest that RD^{Rio} strains may cause a form of TB with a distinct clinical presentation. Overall, the high prevalence of this genotype may be related to enhanced virulence, transmissibility, and/or specific adaptation to a Euro-Latin American host population. The identification of RD^{Rio} strains outside of Brazil points to the ongoing intercontinental dissemination of this important genotype. Further studies are needed to determine the differential strain-specific features, pathobiology, and worldwide prevalence of RD^{Rio} *M. tuberculosis*.

Tuberculosis (TB) is a preventable and curable infectious disease that nonetheless remains a significant cause of morbidity and mortality in resource-poor nations (14). TB also threatens to reemerge in developed nations as a consequence of increased immigration, its synergy with the human immunodeficiency virus (HIV)/AIDS epidemic, and a deprioritization of TB control efforts in public health policy (64). The principal etiologic agent of human TB is *Mycobacterium tuberculosis*, but other species within the *M. tuberculosis* complex (MTC), such as "*Mycobacterium canettii*" (proposed name), *Mycobacterium africanum*, and *Mycobacterium bovis*, are also known to cause TB in humans (30). Although the MTC is considered a relatively homogenous taxon at the DNA sequence level, an increasing number of species-, lineage-, and strain-specific genetic variations have been revealed by the identification of multicopy repeat elements, targeted interrogations of specific genetic loci, and the comparison of whole-genome sequences of several MTC species and strains (21, 24, 31). These differences have been exploited as markers for epidemiological purposes and/or to assist in the identification of MTC strains to the species level (4, 30, 31). As such, these data have informed our understanding of biodiversity among tubercle bacilli with respect to phylogeny, variations in geographic and host preferences, virulence, transmissibility, acquired antimicrobial resistance, and the ability to induce or evade immune responses (7, 13, 23, 42, 43, 46).

An emerging paradigm of MTC evolutionary biology posits that genetic and biological diversity in MTC species and their sublineages stems primarily from gross genome alterations, such as deletions, insertions, inversions, and duplications, which are collectively known as region-of-difference (RD) loci (13). Errors in DNA replication, movement of mobile genetic elements, or recombination between adjacent homologous DNA fragments with loss of the intervening sequence repre-

* Corresponding author. Mailing address: Cornell University, Joan and Sanford I. Weill Medical College, Department of Medicine, Division of International Medicine and Infectious Diseases, Room A-421, 525 East 68th St., New York, NY 10021. Phone: (212) 746-6316. Fax: (212) 746-8675. E-mail: jlho@med.cornell.edu.

‡ L.C.O.L. and R.C.H. contributed equally to this work.

† Supplemental material for this article may be found at http://jcm.asm.org/.

§ Present address: Clinical Microbiology Service and Department of Pathology, Columbia University Medical Center, New York-Presbyterian Hospital, New York, NY 10032.

▽ Published ahead of print on 26 September 2007.

sent some of the irreversible changes in genetic makeup that likely produce RD loci (30). RD loci ranging in size from single-nucleotide deletions to long sequence polymorphisms (LSPs) up to 19.5 kb (29, 31) have been identified, and phylogenetic maps for the evolution of the entire MTC and several of its sublineages have been constructed based on the sequential stepwise acquisition of LSPs (7, 23, 42). For instance, most non-*M. tuberculosis* MTC species can be differentiated from *M. tuberculosi*s sensu stricto and "*M. canettii*" by the characteristic presence of the RD9 deletion (7). Deletions appear to be particularly well suited as specific markers for defining *M. tuberculosis* lineages (23). It is also recognized that discrete acquired single-nucleotide polymorphisms (SNPs) contribute to phenotypic variation within the MTC (26, 48) and that, like LSPs, SNP informatics is a powerful tool for reconstructing familial interrelationships within the MTC (23, 27).

The epidemiological study of tubercle bacilli in combination with genotyping is important for differentiation of patient isolates, cluster analyses, and contact investigations. *M. tuberculosis* isolates with identical DNA fingerprints are said to be clustered, and the proportion of clustering in a population is thought to reflect the amount of recent transmission (25). Large-scale genotyping efforts, by either IS*6110*-based restriction fragment length polymorphism (RFLP) typing (36, 54, 62), spacer oligonucleotide typing (spoligotyping) (16, 33), variable-number tandem repeats of mycobacterial interspersed repetitive units (MIRU-VNTR) analysis (15, 56), deletion typing, or whole-genome SNP analyses, have greatly improved our understanding of MTC population structure at both a regional and a global level (5–7, 20, 23, 27). Spoligotype patterns, for example, allow the grouping of epidemiologically important strain families, such as the Beijing, Haarlem, S, T, X, East African-Indian (EAI), and Latin American-Mediterranean (LAM) clades (9). Although there is a relative paucity of epidemiological studies conducted in countries with high incidences of TB, the available data suggest that families of closely related strains are common in these areas (59). Studies that segregate *M. tuberculosis* according to clade are important because the recognition of specific *M. tuberculosis* subpopulations and/or clones with unique genetic traits and/or particular clinical disease patterns could advance our understanding of the roles of specific genes in pathogenicity and could impact epidemiologically based strategies for combating TB.

The current study was initiated to apply a recently described genomic deletion-based PCR protocol for the differentiation of MTC isolates (30, 31) to a collection of MTC isolates from Rio de Janeiro, Brazil. In the course of that evaluation, we happened upon a novel LSP, now designated RD^{Rio}, that marks a major sublineage of the LAM spoligotype family and the predominant clonally derived *M. tuberculosis* population causing TB in Rio de Janeiro.

MATERIALS AND METHODS

Study setting. According to the World Health Organization, Brazil ranks 15th on the list of high-burden TB countries and is one of the 22 nations in which 80% of the world's new TB cases occur (65). Brazil reported 86,881 new TB cases in 2004, which translates to an incidence rate of 60 per 10^5 persons (65) The overall prevalence of *M. tuberculosis* infection among adults was 77 per 10^5 persons at this time, and the mortality rate from TB in Brazil was estimated to be 7.8 per 10^5 persons, or 14,355 deaths. Approximately 6.2% of adults with TB are thought to be HIV positive, and an estimated 20% of people living with HIV/AIDS have pulmonary TB (66). Approximately 12% of adult AIDS deaths were due to TB (A. L. Kritski, personal communication). Brazil is one of the few high-TB-burden countries that has the potential to reduce the risk of HIV-related TB, owing to its policy of universal access to highly active antiretroviral therapy (45). It should be noted that Brazil's current information systems fail to capture an estimated 30% of TB cases and that the treatment dropout rate is about 13% (A. L. Kritski, personal communication). Vaccination with *M. bovis* BCG was introduced in Brazil in 1940 and has been systematically administered to all infants since 1976 (53).

Rio de Janeiro is the second largest city in Brazil, with ~5.8 million inhabitants. In 2003, Rio de Janeiro City had ~8,414 TB cases per year and an incidence rate of 105.5 per 10^5 persons, almost twice as high as the rest of the country overall (57). In total, 377 deaths related to TB were registered in 2003, with an incidence of 6.3 per 10^5 persons. Official figures estimate that 8% of adult TB patients in Rio de Janeiro have HIV infections (57). Primary health centers administered by the Rio de Janeiro City Health Department provide free medical treatment to TB patients. The Clementino Fraga Filho University Hospital (UH) of the Federal University of Rio de Janeiro (FURJ), Rio de Janeiro, Brazil, serves a population of ~1 million persons. The City Health Care centers have the primary responsibility for treatment of TB cases. The Mycobacteriology Laboratory (ML) of the UH-FURJ provides mycobacteriology culture support to the UH as well as to the City Health Care centers and performs ~5,000 cultures annually, ~20% of which are positive for growth of *M. tuberculosis*.

Handling of clinical specimens. The staff of the ML of the UH-FURJ has been unchanged since 2001. Four technicians are involved in the culture setup of clinical samples, where the entire sputum sample, following sodium hydroxide decontamination, is divided into three equal volumes and plated onto Löwenstein-Jensen (LJ) agar slants. Using a standardized protocol, the mycobacterial culture is (i) inspected under a magnifying glass, first at 48 to 72 h after plating and then twice weekly for as long as 60 days, and (ii) is either recorded as negative after 60 days, given an exact count for colonies numbering 1 to 19, or scored as +1 (20 to 100 colonies), +2 (100 to 200 colonies), or +3 (>201 colonies [where as many as 500 CFU are enumerable]). The mean of the triplicate is recorded as the colony count for the entire culture.

Data collection. A total of 431 cultures from a repository of mycobacterial cultures (corresponding to 336 patients) were randomly selected by a single technician of the ML, UH-FURJ, from 1,202 positive cultures, covering the period between January 2002 and August 2003, and were subcultured. Among these mycobacterial cultures, 12 mycobacteria other than MTC (MOTT) (from 12 patients) were included in a blinded fashion to challenge the performance of our PCR typing panel. Replicate patient samples were collected from different body sites and/or on different days. The isolates were identified as *M. tuberculosis* or as MOTT based on standard criteria. Thermolysates were sent to the Weill Medical College, where determinations of identity by genetic methods were finalized for 425 of 431 samples (332 patients), comprising 404 *M. tuberculosis*, 12 MOTT, and 9 mixed MTC/MOTT samples.

Clinical, radiological, and laboratory data. A retrospective review of all relevant ML UH-FURJ records and the clinical records of the patients was conducted using a standardized questionnaire. Bacteriology data for 400 *M. tuberculosis* patient isolates included the results of smear staining for acid-fast bacilli (AFB), the quantity of colonies on LJ slants, and sensitivity to anti-TB drugs by the proportion method (10). Clinical data included gender, age, HIV status, homelessness, intravenous drug use, institutional exposure (medical facility or prison), skin reactivity on the purified protein derivative test, clinical symptoms, and site of TB. In addition, radiological data recorded the presence of pulmonary cavitation, the extensiveness of lung infiltrates, mediastinal adenomegaly, and pleural effusion. The team conducting the record review were blinded to the results of *M. tuberculosis* strain genotyping. Clinical data were acquired from 146 patient records (47% of all patients, because only the records from patients cared for at the UH-FURJ, and not those from outside health centers, were available).

Cell culture and thermolysis. All clinical cultures were regrown on LJ slants by standard protocols. When the growth was considered to have attained an optimal biomass, 1 ml of TE buffer (10 mM Tris, 1 mM EDTA [pH 7.5]) was added, and the colonies were detached from the solid medium using a 10-μl loop. The turbid liquid was transferred to a 1.5-ml Eppendorf tube containing six 3-mm-diameter glass beads and was pulse vortexed for 2 min inside a biosafety cabinet. Subsequently, 10 μl of 20-mg/ml proteinase K (Fermentas Inc., Hanover, MD) was added, vortexed, and incubated at 60°C for 4 h, with vortexing every half-hour, until most large clumps were dissolved into the solution. Finally, the tubes were incubated at 80°C for an additional 30 min to deactivate both the proteinase K and any surviving mycobacteria. MTC laboratory strains were grown in 7H9 broth and/or 7H11 agar. To determine the sensitivity of the MTC PCR typing panel when culture thermolysates were used as the source of DNA for PCR,

serial 10-fold dilutions of 3-week-old mycobacterial cultures (of various species) were made in TE buffer. These were then plated onto 7H11 agar for CFU counts, and the remaining samples were thermolysed for subsequent PCR testing. The limits of detection of the protocol were estimated based on the final dilution at which clear and interpretable PCR results were obtained and on the calculated CFU counts per milliliter for that dilution. In these experiments, a thermolysate sample volume of 10 μl was used for PCR.

PCR amplification. A previously described PCR-based strategy was utilized for the genetic identification of MTC isolates to the species level (30, 31). The MTC PCR typing panel protocol targets for amplification a set of eight intra- or interspecific RD loci and controls, in separate but simultaneous reactions. The PCRs either amplified successfully or failed, depending on the presence or absence of target genetic regions in each test strain. When the products are separated by agarose gel electrophoresis, the resulting pattern of products is indicative of the MTC species identity (30, 31). The target gene loci and their primer names, primer sequences, and various amplification product sizes, as well as the programs used to amplify them, are listed in Table S1 in the supplemental material. Primer sequences were either obtained from earlier publications or created as previously described (30). For PCR experiments to bridge the novel deletion, primer pairs were generated to amplify targets upstream and downstream of, and successively outward from, the putative LSP locus in a chromosomal walking approach. When PCR amplicons were produced on either side of the locus, the appropriate forward and reverse primers were combined to amplify across the altered site. Only the primer pairs that successfully generated products bridging the RD^{Rio} LSP are listed in Table S1 in the supplemental material. The general PCR recipe was identical to that used previously (30, 31) except that 1, 2.5, or 10 μl of culture thermolysates was utilized as the source of DNA, depending on the turbidity of the sample (the volume of water was adjusted to give a final PCR volume of 50 μl). PCR amplification from DNA thermolysates was performed using the following programs: program 1b (beginning with a denaturation step of 5 min at 94°C, followed by 45 cycles of 1 min at 94°C, 1 min at 60°C, and 4 min at 72°C, and ending with a final elongation step for 10 min at 72°C) and program 2b (program 1b but with an annealing temperature of 65°C). Initially, program 2b was utilized for the MTC PCR typing panel, but in later experiments program 1b was used. In general, more-robust PCR amplicons were produced as a result of this change in protocol (data not shown), and complete typing results with program 1b were acquired for most of the strains that failed to amplify, or that amplified poorly, using program 2b. Mycobacterial strains that repeatedly failed to produce clear and interpretable MTC PCR typing panel results were excluded from further analysis. PCR for principal-genetic-group (PGG) analysis was performed using program 2b, while PCRs for the RD^{Rio} bridge, TbD1 bridge, 16S rRNA, and *hsp65* were performed using program 1b.

PGG analysis, PCR-RFLP, and DNA sequencing. PGG segregation was accomplished by comparative analysis of commonly observed SNPs in $katG^{203}$, $katG^{463}$, and $gyrA^{95}$ using previously described protocols for RFLP analysis of restriction enzyme-digested PCR products (PCR-RFLP) (22, 31, 52). The distribution of the SNPs in *katG* and *gyrA* indicates that PGG1 strains are ancestral to PGG2 strains, which in turn are ancestral to PGG3 strains. Direct sequencing of PCR fragments was also performed as previously described using combinations of amplification primers and/or primers internal to the target amplicon (see Table S1 in the supplemental material) (31). The Lasergene program (DNAStar Inc., Madison, WI) was used to analyze the derived sequence data and to compare these data with DNA sequence information downloaded from online databases. Overall, the combined sequence data from each interrogated PCR product spanned nearly the entire locus at least twice over. Consensus sequences were constructed based on alignments of the sequencing results and a careful examination of each electropherogram trace representation of the data. To date, *M. tuberculosis* sensu stricto strains are more diverse and segregate to PGG1b, PGG2, and PGG3, while other MTC members are restricted to PGG1a or PGG1b (31). The species identity of each MOTT strain was determined by 16S rRNA (>1,500-bp) and *hsp65* (441-bp) sequence analyses queried against the GenBank database using the BLASTN program and our own in-house database of sequences from species-validated strains of MOTT. A similarity of >99.5% to the closest relative 16S rRNA or *hsp65* sequence, along with identity agreement between these loci, was used as the criterion for the determination of species (12, 41).

Spoligotyping. All *M. tuberculosis* strains were spoligotyped using a commercial kit (Isogen Bioscience BV, Maarssen, The Netherlands) (33). The results were recorded both in a 43-digit binary format representing the 43 spacers and as an octal code (9). The person who performed the spoligotyping was blinded to the other results of *M. tuberculosis* strain genotyping. The spoligotype patterns were compared with an updated SpolDB4 database of the Pasteur Institute of Guadeloupe (the initial version is available at http://www.pasteur-guadeloupe.fr:8081/SITVITDemo) that provides information on the shared-type distributions of *M. tuberculosis* spoligotypes worldwide (N. Rastogi and T. Zozio, personal communication). The updated "in-house" version of the Pasteur Institute of Guadeloupe database, at the time of this comparison (November 2006), contained a total of 2,808 shared types corresponding to 63,473 clinical isolates from 121 countries of isolation and 160 countries of origin. Major phylogenetic clades were assigned according to signatures provided in SpolDB4 (for a description, see reference 9).

MIRU-VNTR typing. MIRU-VNTR loci 2, 4, 10, 16, 20, 23, 24, 26, 27, 31, 39, and 40 were individually amplified and analyzed as previously described (56), and the results from each of the 12 loci were combined to form a 12-digit allele profile. At least one isolate from every patient in the RD^{Rio} and wild-type (WT) LAM groups was evaluated.

MST. Spoligotyping and MIRU-VNTR patterns were introduced into Bionumerics (Applied Maths, Kortrijk, Belgium) using an Open Database Connectivity protocol, and a similarity matrix was constructed using, respectively, the Dice and the categorical index. Minimum spanning trees (MST) were also built from spoligotype and MIRU-VNTR patterns separately for a better understanding of the localization of the RD^{Rio} LSP within the strain set analyzed. For this purpose, either a binary (spoligotype patterns) or a categorical (MIRU-VNTR patterns) similarity coefficient was used, and in both cases, priority for linking was given to spoligotype or MIRU-VNTR patterns that had the highest number of single-locus variants (SLVs). The loss and acquisition of single or multiple copies of repeats in different MIRU-VNTR loci were considered equally probable events for MST. The spoligotype MST was built from a core of 37 different LAM spoligotypes obtained from 150 *M. tuberculosis* patient strains. For its construction, a binary similarity coefficient and the highest number of SLVs as a priority rule were used; creation of hypothetical links was allowed. Spoligotypes were excluded from the analysis if there were no intermediate patterns or if they differed in more than 2 spacers from the nearest pattern (originally 67 patterns from 180 LAM patient strains). The MIRU-VNTR MST was built from 113 different MIRU-VNTR patterns obtained from 174 LAM *M. tuberculosis* patient strains. For its construction, a categorical similarity coefficient and the highest number of SLVs as a priority rule were used; creation of hypothetical links was allowed. MIRU-VNTR types were excluded from the analysis if they differed at more than 2 MIRU-VNTR loci from the nearest pattern (originally 119 different patterns from 180 LAM strains).

IS*6110*-RFLP typing and phylogenetic analysis. Chromosomal DNA extraction and IS*6110* RFLP genotyping were performed in accordance with a standard protocol described previously (37, 58), using an IS*6110* right-side-specific probe. The hybridization patterns of clinical *M. tuberculosis* strains were compared on a Sun Sparc 5 workstation (Sun Microsystems) with Whole Band Analyzer software (version 3.5; BioImage). The similarity of hybridization patterns was calculated by Dice coefficient as previously described (28). The phylogenetic tree was constructed using Bionumerics software. Clusters were defined as at least two *M. tuberculosis* strains with identical RFLP profiles isolated from different patients. The investigators performing the methodology were blinded to the WT/RD^{Rio} categorical genotypes of the test strains.

Statistical analysis. Statistical analyses were performed using the STATA 7.0 (Stata Corporation, College Station, TX) and Prism 4 (GraphPad Software, Inc., San Diego, CA) programs. Categorical variables were compared by a two-tailed Fisher exact test. The Mann-Whitney test was used for continuous variables. Multiple logistic regression was employed to infer independent associations between clinical and epidemiological variables and the *M. tuberculosis* RD^{Rio} genotype. Variables were selected for logistic regression when they presented with a P value of <0.15 in univariate analysis. A P value of ≤ 0.05 was considered significant.

Approval of the study. The study protocol was approved by the institutional review boards of FURJ and the Weill Medical College of Cornell University, New York, NY.

Additional Web addresses. Additional Web addresses used were http://www.ncbi.nlm.nih.gov for GenBank and BLAST, http://www.sanger.ac.uk/Projects/M_tuberculosis for the Sanger Center, http://genolist.pasteur.fr/TubercuList for Tuberculist, and http://www.pasteur-guadeloupe.fr/tb/SpolDB4 for SpolDB4.

RESULTS

MTC PCR typing panel results. In Brazil, the proportion of infections caused by MTC species other than *M. tuberculosis* contributing to the overall TB burden has not been well characterized because, in most diagnostic laboratories, the bio-

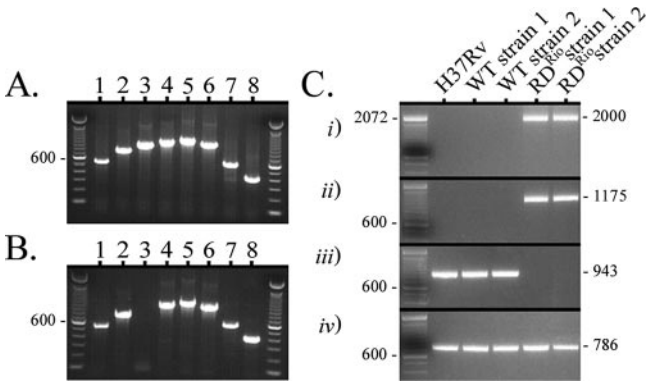

FIG. 1. PCR for RD loci in MTC strains from Rio de Janeiro. (A) MTC PCR typing panel result for a typical WT *M. tuberculosis* strain, showing an uninterrupted arcing pattern. (B) MTC PCR typing panel result for 30% of *M. tuberculosis* strains in the sample set, showing a failure to amplify from the IS*1561′* locus. Lanes: 1, 16S rRNA; 2, *cfp32* (Rv0577); 3, IS*1561′*; 4, Rv1510 (RD4); 5, Rv1970 (RD7); 6, Rv3877 to Rv3878 (RD1); 7, Rv2073c (RD9); 8, Rv3120 (RD12). Unlabeled lanes contain the 100-bp ladder. (C) Composite panel illustrating the representative results of PCR amplification to span the IS*1561′*-deleted (RD^{Rio}) locus (panels i and ii) (bridge PCR 1 and bridge PCR 2, respectively), to target the intact locus (IS*1561′*) (panel iii), and to provide a control amplification (*cfp32*) (panel iv) (30, 31). The first lane contains a marker (100-bp ladder). H37Rv, *M. tuberculosis* strain H37Rv; WT strains 1 and 2, two representative *M. tuberculosis* WT strains; RD^{Rio} strains 1 and 2, two representative *M. tuberculosis* RD^{Rio} strains. Sizes of PCR products (in base pairs) are given on the right.

chemical tests necessary to separate these closely related species are not performed. We applied an MTC PCR typing protocol (30) to 431 mycobacterial clinical isolates from Rio de Janeiro in order to probe for non-*M. tuberculosis* MTC species in this setting. Of the isolates identified as MTC by the clinical laboratory, 404 (corresponding to 312 patients) produced MTC PCR typing panel patterns indicative of *M. tuberculosis* (31). We did not identify any non-*M. tuberculosis* MTC species in this collection of isolates. However, while 70% of strains (283 strains from 219 patients) exhibited the prototypic *M. tuberculosis* pattern of amplicons (Fig. 1A), the remaining 30% of strains (121 strains from 93 patients) consistently and unexpectedly failed to amplify from the IS*1561′* locus (Fig. 1B). Previously, *M. tuberculosis* strains have been identified that failed to amplify from certain loci in the MTC PCR typing panel, and it has been our experience that a novel LSP was present (31). (The summary data from experiments to estimate the sensitivity of our MTC PCR typing panel are presented in Table S2 in the supplemental material and include additional online text related to samples containing MOTT.)

Characterization of the RD^{Rio} deletion. To delineate the boundaries of the potential IS*1561′*-centric LSP, a bidirectional chromosomal walking approach was taken. Using three IS*1561′* PCR-negative Rio de Janeiro *M. tuberculosis* strains as prototypes, PCR products were eventually generated both upstream and downstream of IS*1561′* in the respective PE_PGRS50 (Rv3345c) and *folD* (Rv3356c) genes of each test lysate. The upstream forward primer was then combined with the downstream reverse primer to successfully bridge the altered locus, producing a single 2,000-bp PCR amplicon (Fig. 1C). The products were then sequenced in order to characterize the nature of gene loss at this locus. In later experiments, primers for the bridge PCR were reconfigured so as to amplify a shorter chromosomal fragment (1,175 bp) that also crossed the deleted region. These primers were validated in a set of previously tested isolates and thereafter were used for sequence analysis and to complete the identification of strains.

Sequence analysis of the locus-bridging PCR product provided evidence of a novel RD locus, which we have named RD^{Rio}. In total, 26.317 kb of contiguous DNA sequence was absent (~0.6% of the entire genome) relative to the published sequence of *M. tuberculosis* strain H37Rv (GenBank accession number AL123456). To our knowledge, the RD^{Rio} LSP represents the single longest chromosomal deletion identified in *M. tuberculosis* to date; the largest RD previously reported was ~19 kb (29). Ten genes were either removed (Rv3347c to Rv3354) or modified (Rv3346c and Rv3355c) (Fig. 2) (the locations and descriptions of the RD^{Rio}-associated genes are listed in Table S3 in the supplemental material). Among the genes deleted are two PPE genes (Rv3347c and Rv3350c), which code for hypothetical proteins included in a large family postulated to be expressed on the extracellular surface and considered potential antigens for host immunity (13). The Rv3346c and Rv3355c genes flank those that were deleted in RD^{Rio} and were consequently altered in the production of this LSP. Both genes code for putative integral membrane proteins of unknown function; they are short homologs of one another, have an overall similarity of 85% at the nucleotide sequence level, and share a central stretch of 158 bp that is 100% identical. Sequence analysis of the RD^{Rio} locus failed to identify an exact breakpoint to delineate the boundaries of DNA loss (which would indicate DNA polymerase errors as the mechanism of deletion) but rather uncovered an in-frame chimeric gene in place of the deleted sequence. The chimera has Rv3346c proximally/5′ (in relation to the plus strand of the mycobacterial chromosome) and then transitions across the 158-bp zone of perfect homology into Rv3355c distally/3′ (see Fig. S1A and B in the supplemental material). On the basis of these data, homologous recombination between Rv3346c and Rv3355c as a result of a double-crossover event, with excision and loss of the intervening genomic segment, is the proposed mechanism for the derivation of the RD^{Rio} locus and the amalgamation of Rv3346c and Rv3355c to form the new gene Rv3346/55c (see Fig. S1C in the supplemental material). No tandem repeat sequences were evident at the site of gene loss to indicate that the movement of mobile transposable elements may have contributed to the mechanism of deletion (37).

PCR experiments were then conducted to evaluate the distribution of the RD^{Rio} LSP within the collection of *M. tuberculosis* strains from Rio de Janeiro ($n = 404$). All IS*1561′* PCR-negative *M. tuberculosis* strains ($n = 121$) amplified an RD^{Rio} bridge PCR product, while most of the IS*1561′* PCR-positive, or "wild type," *M. tuberculosis* strains ($n = 279$) failed to generate an RD^{Rio} bridge PCR band (Fig. 1C). WT strains failed to amplify presumably because the size of the predicted PCR product, at more than 26.3 kb, is prohibitive under the PCR conditions that were utilized. Four samples from three patients, in contrast, repeatedly produced both IS*1561′* and RD^{Rio} bridge PCR products, suggesting a mixed sample population of WT and RD^{Rio} *M. tuberculosis* strains. To validate

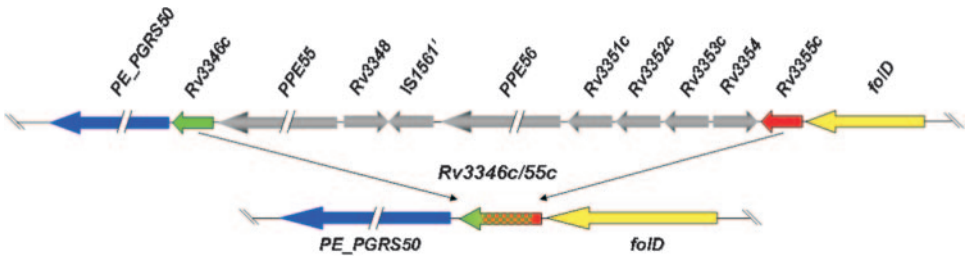

FIG. 2. Diagram showing the genes deleted or altered in the derivation of the RD^{Rio} LSP. The recombination between the homologous pair Rv3346c and Rv3355c formed a new in-frame fusion gene (Rv3346c/55c), and all elements between these two genes were excised, resulting in a deletion of 26.314 kb of contiguous DNA sequence.

this hypothesis, PCR-confirmed WT and RD^{Rio} clones were isolated from each of the two sputum cultures (provided on different days) from patient 315. These four clones were included in the subsequent molecular fingerprint analyses as strains 315c and 315e (WT) and strains 315d and 315f (RD^{Rio}). No attempts were made to obtain clones from the other two mixed WT/RD^{Rio} samples (each from a different patient), because laboratory contamination could not be conclusively ruled out based on a single culture, and these were excluded from further analysis. The RD^{Rio} bridge PCR products from all strains positive, including 2 of the clonal strains ($n = 123$), were sequenced and found to be identical across the Rv3346/55c chimera.

PGG and TbD1 analysis of Rio de Janeiro *M. tuberculosis* strains. The RD^{Rio} LSP molecular data suggested that RD^{Rio} strains are clonal in origin. However, the mechanism of deletion for the RD^{Rio} LSP, i.e., homologous recombination between nonmotile genetic elements, may theoretically have been independently reproduced in different lineages as homoplastic LSPs. For this reason, data from additional molecular markers were evaluated in order to clarify the genetic relationships among RD^{Rio} LSP-bearing *M. tuberculosis* strains. Further molecular characterization of the entire collection of Rio de Janeiro *M. tuberculosis* strains began with PGG analysis (22, 31, 52). Notably, the *M. tuberculosis* WT strains were represented in all three PGGs, whereas the *M. tuberculosis* RD^{Rio} strains belonged exclusively to PGG2, a result that supports the hypothesis that the latter strains are linked. Table S4 in the supplemental material lists the acquired molecular data by patient and isolate. Lastly, a subset of WT ($n = 20$) and RD^{Rio} ($n = 16$) strains, including the four clonal strains, was evaluated for the presence or absence of the TbD1 deletion (31). All *M. tuberculosis* RD^{Rio} strains had the TbD1 deletion ("modern" lineage) (7).

Spoligotyping of Rio de Janeiro *M. tuberculosis* strains. To further dissect the genetic population structure of *M. tuberculosis* in Rio de Janeiro, and of the RD^{Rio} strains in particular, we spoligotyped the collection of tubercle bacilli and clonal strains (404 isolates from 310 patients). The results of spoligotyping are provided by patient in Table S4 and summarized in Table S5, both in the supplemental material, and are described in greater detail in the accompanying online text. Of particular interest was the observation that all RD^{Rio} strains belonged exclusively to the LAM family ($P \leq 0.0001$ by a two-sided Fisher exact test). The LAM spoligotype signature is characterized by the simultaneous absence of at least spacers 21 to 24 and 33 to 36, and the LAM family is one of the most prevalent and widely distributed strain families in the world (8, 9). Indeed, 58% of all Rio de Janeiro spoligotype patterns were LAM types. LAM9 is the basic LAM spoligotype and presumably the originator of the LAM lineage (9). The RD^{Rio} patient strains (34 different spoligotypes) were subtyped mainly as LAM9 (23%), LAM1 (16%), or LAM2 (17%); a smaller number were LAM4 (2%), LAM5 (3%), or LAM6 (2%); and the remaining 37% were a mixture of LAM derivative spoligotypes (50% LAM9-like, 18% LAM1-like, and 32% LAM2-like). On the other hand, LAM3 (19%) was the predominant type among the WT LAM strains; a smaller number were LAM6 (13%), LAM5 (5%), LAM9 (5%), or LAM4 (2%); and the remaining were various other LAM derivative types (56%). Therefore, LAM1 and LAM2 were exclusively of the RD^{Rio} genotype, LAM3 was solely represented in the WT genotype, and LAM4, LAM5, LAM6, and LAM9 were represented in both the RD^{Rio} and WT genotypes.

To visualize the potential phylogenetic interrelationships among LAM strains in the sample set, an MST (Fig. 3) was created using the main LAM spoligotypes and the derivative spoligotypes that were part of this complex (37 spoligotypes from 150 strains). As a result, the RD^{Rio} LSP indeed appears to have originated in a WT LAM9 progenitor that then diversified and expanded with the successive loss or disruption of direct-repeat spacer elements. It is quite intriguing that patients with RD^{Rio} LAM9 strains outnumbered patients with WT LAM9 strains by a factor of 4.2, a finding that hints at a difference in fitness. The existence of both WT and RD^{Rio} LAM4, LAM5, and LAM6 strains is suggestive of spoligotype convergence but may also be due to RD^{Rio} LSP homoplasy, as mentioned above. Lastly, the majority of LAM strains are known to be included in genetic cluster VI in whole-genome SNP analyses (27, 34); however, it remains to be confirmed whether the same can be said for RD^{Rio} strains.

MIRU-VNTR typing of Rio de Janeiro *M. tuberculosis* LAM strains. In order to more fully characterize the RD^{Rio} strains and to differentiate them from WT LAM strains, MIRU-VNTR typing was performed on LAM isolates of the Rio de Janeiro collection. At least one isolate from each patient ($n = 180$ TB cases), and each of the patient 315 clonal strains (including here the non-LAM WT clone as well) ($n = 4$), was evaluated (189 strains in total; MIRU-VNTR typing was performed twice for five patients). Importantly, when multiple strains from the same patient were evaluated by MIRU-VNTR, the results were concordant, as were those for each of

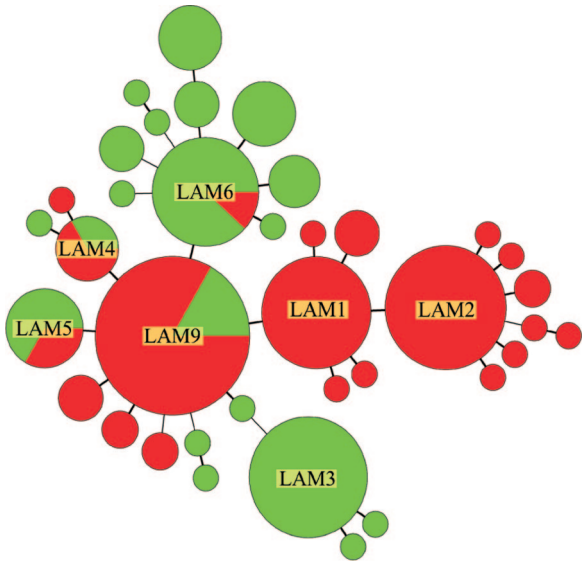

FIG. 3. An MST of spoligotype data from Rio de Janeiro *M. tuberculosis* LAM strains is consistent with the parsimonious phylogeny of WT and RD^{Rio} genotype strains. Colors indicate RD^{Rio} (red) or WT (green) strains. Each circle represents a particular spoligotype, and the size of a circle is relative to the number of strains with that spoligotype. Thicker lines linking circles indicate a 1-spacer difference, and thinner lines indicate a 2-spacer difference, between the spoligotype patterns.

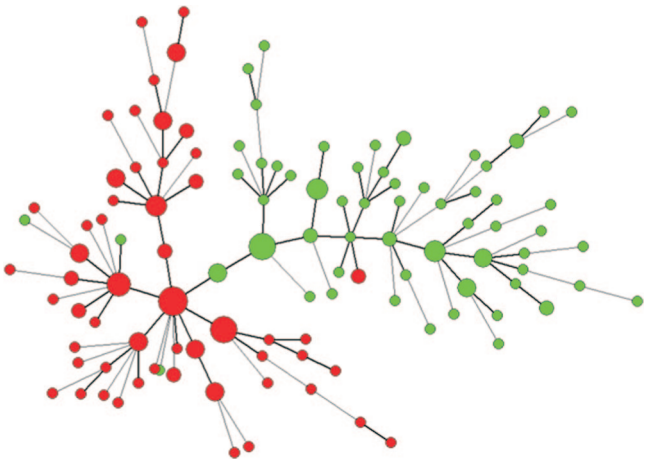

FIG. 4. An MST of MIRU-VNTR data from Rio de Janeiro *M. tuberculosis* LAM strains illustrates the genotypic segregation of WT and RD^{Rio} genotype strains. Colors indicate RD^{Rio} (red) or WT (green) strains. The size of each circle is relative to the number of strains with a particular MIRU-VNTR type. Each circle represents a single MIRU-VNTR profile. Thicker lines linking circles indicate a single-locus difference, and thinner lines indicate a 2-locus difference, between the profiles. The central RD^{Rio} node represents strains with the hypothetical founding RD^{Rio} MIRU-VNTR allele (224226153321) (see Table S7 in the supplemental material).

the pairs of patient 315 clonal strains. The results of MIRU-VNTR typing are provided in Table S4 and summarized in Table S6, both in the supplemental material, and are described in the accompanying online text. There were 119 different MIRU-VNTR profiles among the LAM spoligotype strains. When these strains were segregated by RD^{Rio} or WT genotype, 61% of RD^{Rio} patient strains exhibited shared patterns (2 to 7 patients per cluster) while 38% of WT LAM patient strains demonstrated shared patterns (2 to 6 patients per cluster). The tighter association of RD^{Rio} strains into genotypic clusters by MIRU-VNTR ($P = 0.0029$ by a two-sided Fisher exact test) indicates that as a whole, RD^{Rio} strains may be a more significant source of recently transmitted TB than *M. tuberculosis* WT strains in Rio de Janeiro. However, a larger population-based epidemiological study is required in order to confirm this hypothesis.

The allelic diversity at each MIRU-VNTR locus, segregated by RD^{Rio} and WT LAM populations, was then determined. As shown in Table S7 in the supplemental material and described in greater detail in the accompanying online text, RD^{Rio} and WT LAM strains can be differentiated by MIRU2 and MIRU40. In fact, MIRU40 was highly discriminatory for *M. tuberculosis* RD^{Rio} patient strains, since 100% of these strains ($n = 94$) and just 2% of *M. tuberculosis* WT LAM patient strains (2 of 86) had a single unit at this locus. MIRU40 has been shown previously to exhibit the greatest global allelic diversity and to be the most differentiating locus for LAM strains (19, 51). The MIRU2 allele copy number was also highly indicative of *M. tuberculosis* RD^{Rio} patient strains, since 98% (91 of 94) of these strains had two unit copies at this locus. Indeed, among the LAM family tubercle bacilli evaluated, the combination of a MIRU2 copy number of 2 and a MIRU40 copy number of 1 was a signature restricted to 98% of RD^{Rio} strains; this combination may be helpful in the future to identify additional isolates within databases listing MIRU-VNTR-typed LAM strains. On the basis of the MIRU-VNTR results, the RD^{Rio} and WT strains that shared the LAM4, LAM5, LAM6, or LAM9 spoligotype were clearly distinguished from each other, arguing that spoligotype convergence was masking the actual phylogenetic separation of these strains.

To illustrate the genotype-based segregation of the RD^{Rio} and WT LAM groups by MIRU-VNTR, an MST was constructed from 113 different MIRU-VNTR patterns representing 174 *M. tuberculosis* LAM strains (Fig. 4). The MST segregated the MIRU-VNTR patterns from WT LAM strains into a separate offshoot from the RD^{Rio} isolates. There were, however, five exceptions: three WT LAM strains (from patients 110, 112, and 151) that were located in the RD^{Rio} group and two RD^{Rio} strains (patients 69 and 70) that were linked to the WT LAM radiation. These misclassifications are attributed to limitations in the computer program algorithm, because the MIRU-VNTR patterns of the WT LAM outliers at MIRU2 and MIRU40 clearly placed them in the correct genotypic category, while the RD^{Rio} outlier strains were those that uniquely possessed 1 copy of the MIRU2 allele. The MST and overall MIRU-VNTR typing results clearly supported the existence of a single, now diversified RD^{Rio} lineage that has expanded from a common progenitor, even though the MIRU-VNTR analysis did not completely correspond with the phylogenetic structure of the MST generated by the spoligotype patterns. This is an increasingly recognized limitation of the current 12-locus MIRU-VNTR system when diverse *M. tuberculosis* collections are considered (55).

IS*6110*-RFLP typing of Rio de Janeiro *M. tuberculosis* strains. To gain perspective on the breadth of IS*6110*-RFLP

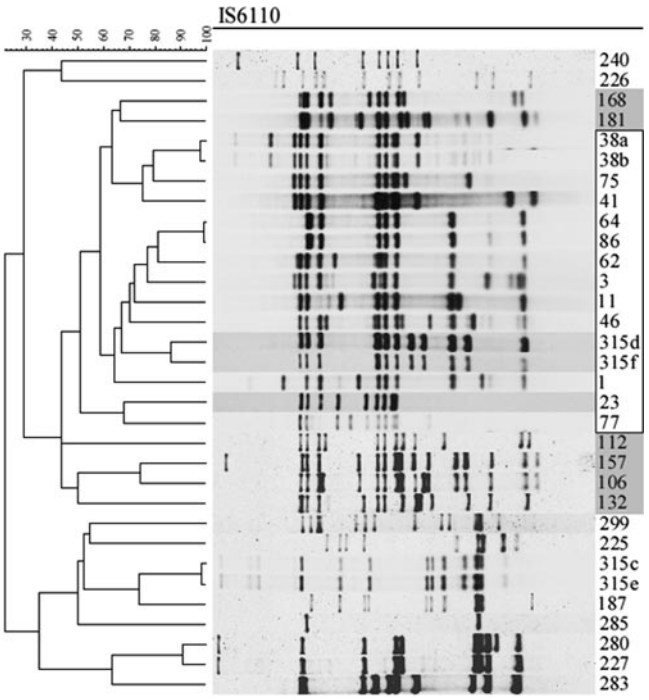

FIG. 5. *M. tuberculosis* RD^{Rio} strains group together in phylogenetic analysis of IS*6110*-RFLP fingerprint data. IS*6110* Southern blot hybridization and dendrogram construction were performed as described in Materials and Methods using a set of available Rio de Janeiro *M. tuberculosis* DNA samples. Patient strain numbers are given on the right. RD^{Rio} patient strain numbers are lightly shaded and boxed; WT LAM patient strain numbers have dark shading; and WT non-LAM patient strain numbers are neither shaded nor boxed. Note that the clonal strains in each separate pair from patient 315 are identical to each other, as are the two strains from patient 38. Interestingly, the IS*6110*-RFLP types from patient strains 64 and 86 were indistinguishable, and their MIRU-VNTR types were identical, but their spoligotypes differed by 1 spacer (see Table S4 in the supplemental material), suggesting that either they may not in actuality form a directly linked cluster or they may represent an example of strains evolving during transmission. The phylogenetic tree was constructed using Bionumerics software, and the similarity of hybridization patterns was calculated by the Dice coefficient; the Pearson correlation Opt value was 4.00% (range, 0.0% to 100%).

types of *M. tuberculosis* RD^{Rio} strains relative to WT strains in Rio de Janeiro, we evaluated a subset of available isolates (32 in total, comprising 15 RD^{Rio}, 6 WT LAM, and 11 WT non-LAM isolates) that had been recovered from 29 patients and that included each of the patient 315 clonal strains. The overall fingerprint patterns of LAM strains possessed 8 to 13 IS*6110* bands and were clearly distinguished from those of non-LAM strains by phylogenetic analysis (Fig. 5). Most notably, the RD^{Rio} strains formed a consolidated subgroup among the LAM isolates. However, there were no hybridization bands at a specific position that distinguished *M. tuberculosis* RD^{Rio} strains from other LAM strains. Clearly, an expanded comparative examination of the RFLP types of Brazilian WT LAM and RD^{Rio} genotype *M. tuberculosis* strains is needed in order to better define the differences between the groups. Nonetheless, all LAM strains appeared to be related to the 8-band pattern (FO, according to the Public Health Research Institute TB Center *M. tuberculosis* RFLP genotype nomenclature) exhibited by patient strains 64 and 86 (see Table S4 in the supplemental material) and shared two 2- to 3-band sets running at about 5.5 to 7.0 and 2.5 to 3.5 kb, respectively. This particular 8-band pattern is very similar to an RFLP type previously reported to be the most frequently observed type in Rio de Janeiro and Rio Grande do Sul, Brazil; is nearly identical to that of a widespread strain in São Paulo, Brazil; and is known to be present in databases of isolates originating from other South American, Caribbean, European, and African countries (2, 53). As reported by Suffys et al. (53), the 8-band FO-like pattern was related to more than half of Brazilian RFLP types. Combined, these data support the idea that the LAM family is an important cause of TB worldwide (9) and further raise the possibility that international transmission of *M. tuberculosis* RD^{Rio} strains to and from Brazil could be occurring frequently.

International spread of *M. tuberculosis* RD^{Rio} strains. In a previous study, we identified four PGG2 TbD1-deleted *M. tuberculosis* strains that failed to amplify from the IS*1561′* locus, but the underlying basis of PCR failure was not determined (31). These strains were isolated from patients living in The Netherlands or Djibouti. Outside the aforementioned study, we have identified an additional three strains from the United States that failed to amplify from the IS*1561′* locus. More-detailed information on these strains is published as supporting online text (supplemental material). Since we recognized that the IS*1561′*-deleted strains isolated outside of Brazil may be members of the RD^{Rio} lineage, they were genotypically evaluated similarly to the Rio de Janeiro RD^{Rio} strains (see Table S4 in the supplemental material). Each produced an RD^{Rio} bridge PCR amplicon of the expected size, belonged to PGG2, had the TbD1 deletion, showed a LAM spoligotype signature, and had a MIRU-VNTR profile with 2 copies of the MIRU2 allele and 1 copy of the MIRU40 allele. Hence, every one of the international strains possessed each of the defining genotypic criteria of an *M. tuberculosis* RD^{Rio} strain. Notably, the spoligotype and MIRU-VNTR profiles of each epidemiologically linked pair were indistinguishable, and the MIRU-VNTR types of several of the international strains were either identical or very similar to the proposed MIRU-VNTR profile of the RD^{Rio} progenitor strain (see Table S7 in the supplemental material). Indistinguishable IS*6110*-RFLP profiles were found for the epidemiologically linked strain pairs from a zookeeper and bonobo monkey as well as from a mother and son (data not shown). The IS*6110*-RFLP patterns of all six strains isolated in The Netherlands and the United States were also similar to the FO 8-band pattern (data not shown). Although the number of international RD^{Rio} strains in this study is limited, these data provide evidence of direct transmission of RD^{Rio} strains and of the potential of the RD^{Rio} genotype to cause primary progressive TB and to evolve a multidrug-resistant phenotype. These data also point to the ongoing dissemination of *M. tuberculosis* RD^{Rio} strains outside of Brazil, elevating in importance the potential contribution of this lineage to the worldwide TB burden.

Microbiological quantitation and retrospective patient record reviews. Previous studies have reported that TB caused by the Beijing lineage was associated with multidrug resistance, coinfection with HIV, homelessness, and institutional (e.g., hospital or prison) transmission (11, 17, 32). To evaluate for

risk factors, we performed a retrospective review of the records of 146 patients (those who received care at the UH-FURJ, representing 47% of all patients) and later linked the outcome data to the determined *M. tuberculosis* RD^{Rio}/WT genotypic identity of the infecting strain (this patient group comprised 49% of all patients infected with RD^{Rio} strains [$n = 46$] and 46% of all WT strain-infected patients [$n = 100$]). Compared to patients infected with WT strains, patients infected with *M. tuberculosis* RD^{Rio} strains showed similar rates of drug resistance and HIV-1 coinfection (see Table S8A and B in the supplemental material). The similar percentages of HIV coinfection for WT and RD^{Rio} TB cases (28% and 26%, respectively) are higher than the reported national rate of ~6.2% (66) and reflect the practice of referral of TB cases with comorbid conditions for treatment at the UH-FURJ. Homelessness and institutional transmission did not appear to be risk factors for *M. tuberculosis* RD^{Rio} infection (data not shown). We also did not identify regional clustering within Rio de Janeiro when residence locations were plotted onto a map of the city (see Fig. S2 in the supplemental material), or any temporal clustering of RD^{Rio} strains during the 20 months from which isolates were chosen (data not shown). Patients infected with *M. tuberculosis* RD^{Rio} or WT strains presented to physicians with similar durations of illness, yet patients infected with *M. tuberculosis* RD^{Rio} strains manifested hemoptysis (the coughing of blood) more frequently ($P = 0.05$ by a two-sided Fisher exact test) (see Table S8B in the supplemental material). A higher proportion of patients infected by *M. tuberculosis* RD^{Rio} strains also recorded a history of weight loss ($P = 0.09$ by a two-sided Fisher exact test). When only pulmonary cases were considered, weight loss was significantly associated with *M. tuberculosis* RD^{Rio} infection ($P = 0.04$ by a two-sided Fisher exact test). Available microbiological records of all cultures with recorded colony counts and AFB-positive smears by microscopy were analyzed (see Table S8A in the supplemental material). TB caused by *M. tuberculosis* RD^{Rio} strains was associated with higher bacillary loads (>200 colonies) than that caused by WT strains ($P = 0.04$ by a two-sided Fisher exact test). In contrast, qualitative AFB smear results were similar for RD^{Rio} and WT strains. We next restricted the clinical analyses to just pulmonary TB cases and performed multiple logistic regression analyses for hemoptysis and weight loss; age and sex were included as potential confounders. Following multivariate analyses, weight loss remained significantly associated with TB caused by *M. tuberculosis* RD^{Rio} strains ($P = 0.005$; odds ratio, 6.59; 95% confidence interval, 1.74 to 24.96) (see Table S8C in the supplemental material). Neither sex nor age was a risk factor for infection by an *M. tuberculosis* RD^{Rio} strain.

DISCUSSION

We have uncovered a novel LSP that unifies a host of diverse, heretofore unlinked strains into a single major clade of the LAM family. The *M. tuberculosis* RD^{Rio} strains all belonged to PGG2, showed a LAM spoligotype, and shared a MIRU40 allele copy number of 1 in their MIRU-VNTR profiles, and those that were evaluated had the TbD1 deletion and a related 8- to 13-band IS*6110*-RFLP fingerprint. These genotyping data combine to support the hypothesis that the RD^{Rio} LSP was derived from a unique-event polymorphism that was thereafter inherited by descent from a single mother clone. Of potentially greatest importance, the RD^{Rio} genotype appears to be the predominant single phylogenetically linked *M. tuberculosis* strain causing TB in Rio de Janeiro, Brazil.

The RD^{Rio} LSP is itself quite interesting in that it provides a rare example of gene deletion in *M. tuberculosis* occurring as a result of homologous recombination between nonmobile protein-coding genes a great distance apart (>26.3 kb). Homologous recombination as a mechanism of deletion in *M. tuberculosis* has been described chiefly as an occurrence between adjacent IS*6110* elements or multicopy tandem repeat elements (18). Indeed, it has been established previously that allelic exchange in *M. tuberculosis* is more efficient when a long DNA fragment (>4 kb) is used as the recombination substrate (18). The thought is that short DNA sequences do not efficiently induce recombination because of infrequent positioning and because a minimal degree of overlapping homologous DNA sequence may be required, i.e., ~1.3 kb (the length of IS*6110*), in order to establish an energetically stable association that will spur either spontaneous or RecA-mediated recombination (18). Given the short region of perfect homology (158 bp) between Rv3346c and Rv3355c, as well as the significant distance between these genes, the RD^{Rio} LSP is likely a rare chance product. Homologous recombination between nonmobile genes therefore provides a new mechanism for generating *M. tuberculosis* variants with potentially advantageous phenotypes (40).

The diversity of spoligotype families in Rio de Janeiro that we observed in this study is consistent with a racially diverse and mixed open society with multiple introductions of competing clonal types from abroad. The majority of isolates were of the modern PGG2 or PGG3 Euro-American major phylogeographical lineage (23), consisting primarily of the LAM, Haarlem, and T spoligotype clades—the major genotypic families most frequently found in Africa, Europe, Central America, and South America (9). A detailed look at the SITVIT database (see Materials and Methods) revealed that T and T-like, LAM and LAM-like, and Haarlem and Haarlem-like spoligotypes account for 16.7%, 15%, and 12.1% of *M. tuberculosis* isolates worldwide, respectively (N. Rastogi, personal communication). Although most often noted, and thus far the focus of greatest research interest, the Beijing and Beijing-like spoligotypes account for just 10.3% of *M. tuberculosis* strains globally. Because the T clade classification includes a number of phylogenetically distinct subgroups, the LAM lineage (at 15%) may well be the single *M. tuberculosis* genetic family causing the greatest proportion of TB cases in the world. Whereas the Beijing, Haarlem, and EAI families have been validated by other investigators (19), the LAM family has remained less well defined as a phylogenetic clade (50) and is disproportionately understudied relative to its potential importance. Indeed, as many as 50% of strains in South America are estimated to belong to the LAM family (9), while LAM strains were responsible for 58% of infections in Rio de Janeiro in the sampling described here. Hopefully, studies such as this will raise the profile of the LAM family and make it a focus of greater research attention.

The diversity of existing fingerprint signatures of LAM strains is also one possible reason why the TB epidemic caused

by *M. tuberculosis* RD^{Rio} strains in Rio de Janeiro has gone unrecognized until now. Although the basic Beijing spoligotype (ST-1; absence of spacers 1 to 34) is the signature most often described and is the highest-ranked single shared type in the SpolDB4 international spoligotyping database, several LAM signatures are also among the most highly represented spoligotypes; LAM9 (ST-42) is ranked 4th after T1 (ST-53) and Haarlem3 (ST-50), while LAM1 (ST-20) ranks 8th, LAM3 (ST-33) ranks 12th, and LAM2 (ST-17) ranks 13th overall among single *M. tuberculosis* spoligotypes in SpolDB4 (9). The identification of the RD^{Rio} LSP allowed us to subdivide the LAM family in a new way, along separate major LAM9→LAM3 and LAM9→LAM1→LAM2 evolutionary tracts. These spoligotypes have also been identified in substantial numbers in other countries of the Americas, as well as in Europe and Africa (9). Because all LAM1 and LAM2 strains in this study were exclusively RD^{Rio} strains, these data suggest that RD^{Rio} strains are not restricted to Brazil but are actually circulating worldwide, possibly in significant numbers, a notion supported by our identification of RD^{Rio} strains from countries outside of Brazil. In addition, epidemiological studies describing TB in Caribbean countries have provided MIRU-VNTR profiles for LAM strains identical to the prototype for an RD^{Rio} isolate at MIRU2 and MIRU40 (51). Clearly, the distribution of *M. tuberculosis* RD^{Rio} strains worldwide, as well as the relative impact of this novel genotype on the global TB burden, needs to be established.

As with the LAM family, Beijing genotype strains are disseminated globally, but because the latter are much more stable in their spoligotype and MIRU-VNTR patterns, they are thought to have expanded rapidly from a recent common ancestor (6, 59). By comparison, the rich diversity in fingerprint patterns of LAM strains in Rio de Janeiro observed in this study leads to the hypothesis that the LAM family has had a relatively long history of circulation in Brazil. Given the prevalence of the LAM family in Southern Europe and around the Mediterranean basin, it has been suggested that the presence of the LAM lineage in South America and the Caribbean may be linked to early post-Columbus European colonial history, particularly that of the Spanish Empire (8). For the RD^{Rio} strains, significant microevolution has occurred in direct-repeat, MIRU-VNTR, and IS*6110*-RFLP patterns as well, indicating that the RD^{Rio} LSP arose as a temporally distant genetic event in an ancestral *M. tuberculosis* clone now endemic in Rio de Janeiro. Whether the original RD^{Rio} strain evolved in Brazil from a WT LAM9 strain and is now disseminating beyond South America or whether it emerged elsewhere and was imported to Brazil independently of other WT LAM *M. tuberculosis* strains remains to be determined. The relatively high prevalence of RD^{Rio} strains in Rio de Janeiro (30%) may stem from inherent genetic or phenotypic properties lending a specific selective advantage and/or may be the result of founder effects, having been introduced early and in the absence of other competing genotypes.

Two of the genes lost in the derivation of the RD^{Rio} LSP were PPE genes (*PPE55* and *PPE56*). Of note, antigenic variation is among the putative functions assigned to PE/PPE protein family members (13). Indeed, PE/PPE genes have been reported to be the main source of variation between *M. tuberculosis* strains and other MTC species in whole-genome comparisons (21) and are known to be overrepresented among the MTC genomic LSPs detected to date (21, 44, 47). Some PPE genes have been shown to be independently deleted or modified in multiple MTC species and lineages of *M. tuberculosis* (31): (i) *PPE55* and *PPE56* have been lost in *Mycobacterium microti* and *Mycobacterium pinnipedii* as a result of the MiD3 LSP and in *M. bovis* strain AF2122/97 via frameshifting (24, 31); (ii) a *PPE55* fragment has been shown to be highly polymorphic in "*M. canettii*" (31); and (iii) the *PPE55* gene is disrupted by an IS*6110* element in a sublineage of the Cameroon *M. tuberculosis* family (31, 44). Moreover, *PPE55* and *PPE56* are known to be induced in *M. tuberculosis* both in vivo, in a murine model of TB (60), and in vitro, when mycobacteria were phagocytosed by gamma interferon-activated-macrophages (61). We therefore speculate that *PPE55* and *PPE56* may be mutated or lost by tubercle bacilli as a mechanism to evade immune recognition and that their deletion may have contributed to the success of the RD^{Rio} genotype. This hypothesis is supported by the recent observation that PPE55 is an immunodominant B-cell target antigen that elicited a serologic reaction in 81 to 97% of (non-Brazilian) patients evaluated in one study (49). As such, *M. tuberculosis* RD^{Rio} infection may be a good model with which to evaluate, in natural human infection, the consequences of the loss of an otherwise immunogenic protein for the response to treatment, the severity of disease, the recurrence of infection, the proportion of dual infections, and mortality. The RD^{Rio} deletion was most likely not detected in prior evaluations using a genome microarray approach because members of the PE/PPE gene family were excluded from the analysis (6). Our findings, together with previously noted polymorphisms (34), underscore the need to advance investigations of the PE/PPE family.

Recent evidence supports the notion that *M. tuberculosis* strains with certain genotypes may be more virulent than others (3, 38). The prevalence of RD^{Rio} strains in the high-incidence TB area that is Rio de Janeiro suggests selective advantages of strains with this genotype in terms of their capacity to establish infection and/or their ability to produce disease. Indeed, RD^{Rio} strains caused a disproportionately large number of genotypically clustered cases by MIRU-VNTR analysis in this study, implying that they may be spreading more effectively than WT strains in Rio de Janeiro. The higher proportion of RD^{Rio} strains with the LAM9 spoligotype (the presumed clonal source of the RD^{Rio} deletion) compared to WT LAM strains also supports an added evolutionary advantage to the RD^{Rio} lineage. The success of these strains could be due to differences in gene expression of virulence factors or in immune evasion and/or may be related to the genetic, cultural, or environmental characteristics of the host population.

The relative virulence of *M. tuberculosis* RD^{Rio} strains is most likely the outcome of a long, complex, and dynamic interaction between host and microbial properties rather than being solely rooted in a single breakthrough genomic deletion. New data indicate that certain well-conserved genotypes seem to prevail in areas with a high incidence of TB because they are adapted to that specific human population and maladapted to others (23). It is believed that the successive loss of DNA in tubercle bacilli facilitates the appearance of more-successful pathogens in certain kinds of hosts (31); *M. bovis* is a prime example of an MTC species downsizing and refining its ge-

nome in order to expand its host range (24). Thus, we speculate that *M. tuberculosis* RD^{Rio} strains may be specifically adapted to a Euro-Latino population of hosts. Studies to address this question are ongoing. In parts of China, as many as 90% of TB cases are caused by the Beijing family (1), and this genotype is rapidly emerging in new host populations as relatively homogeneous clonal types (59). Therefore, the 30% prevalence of RD^{Rio} strains that we observed in this study may reflect a rate that is on the rise and has not yet reached its peak. Alternatively, because Brazil is populated by persons of diverse and admixed ethnic and geographical origins, the 30% prevalence of RD^{Rio} strains (of various subclonotypes) may reflect a state of static equilibrium between *M. tuberculosis* RD^{Rio} strains and susceptible hosts in Rio de Janeiro. Further studies are needed to resolve this issue. Indeed, specific LSPs distinguish the Beijing family and its sublineages and may have had positive adaptive consequences (23, 35). A large gene deletion in *M. tuberculosis* was also recently shown to confer immune evasion through the induction of the anti-inflammatory cytokine interleukin-10 (43). In the case of the RD^{Rio} LSP, the shedding of the two PPE genes may have been a microbial mechanism to adapt to the immunogenetic background of the human host population and thus allow the tubercle bacilli to evade immune recognition.

Simultaneous infection with multiple *M. tuberculosis* strains is known to occur at low frequencies (63), and one appealing theory put forth recently posits that, in mixed-strain infections, the second strain is more virulent, allowing it to evade the immune response brought to bear on the first strain. In this study, we confirmed one case of mixed RD^{Rio} and WT infection and noted an additional two possible cases. Mixed-strain infections have the potential to confuse the interpretation of molecular epidemiological data. However, the dual infections we identified were found by screening isolates for both the RD^{Rio} LSP and the original intact locus; mixed infections that included a Beijing strain were similarly identified by targeting a Beijing-specific RD locus by PCR (63).

Medical-chart reviews did not detect specific risk factors for *M. tuberculosis* RD^{Rio} infection, such as drug resistance, HIV coinfection, homelessness, or exposure within medical institutions or prisons. Despite the finding that patients infected with a WT or RD^{Rio} strain had similar durations of illness prior to presentation, we noted several trends that suggested that infection with an RD^{Rio} strain may cause a "more severe" form of TB. *M. tuberculosis* RD^{Rio}-infected patients had greater hemoptysis, increased weight loss, and higher bacterial loads in their sputa. Higher CFU counts are known to indicate accelerated bacterial multiplication in vivo (38) and may increase the rate of transmission, given that TB cases with positive AFB smears are more infectious than respiratory smear-negative cases (39). Our findings further suggest that transmission of RD^{Rio} strains remains an actively ongoing process. The capacity of RD^{Rio} strains to transmit and cause disease is underscored by two epidemiologically linked transmission pairs in the international collection (i.e., transmission from a zookeeper to a bonobo and from an adult child to his mother). Although important for hypothesis building, the trends noted above should be interpreted with caution, because medical records were available for review in only 46% of WT cases and 49% of RD^{Rio} cases. Nonetheless, the indicators from the limited patient chart reviews do suggest that RD^{Rio} strains may cause a "more severe" form of TB, while the higher sputum colony counts may be a marker of increased transmissibility. However, in contrast to our proposed hypothesis, one could easily suggest several other potential explanations for these data that contraindicate a heightened virulence of RD^{Rio} strains. Of particular note is the lack of a standardized volume for sputum specimens inoculated onto LJ slants for CFU quantitation. However, the volume of sputum may be an indicator of infectiousness either because of greater ease of sputum-productive cough and/or for heightened aerosol generation. Given these issues, a prospective study involving more patients is currently under way to better ascertain the true clinical features and risk factors of infection with *M. tuberculosis* RD^{Rio} strains.

Despite the bias that may have been introduced in this study by the small sample size and reliance on a retrospective clinical and microbiological review, we interpret the genetic data and trends in patient data to suggest that *M. tuberculosis* RD^{Rio} strains may be more pathogenic and possibly more infectious than WT strains. The finding that similar proportions of RD^{Rio} and WT TB patients received care at the UH-FURJ and City Health Clinic centers suggests that no overt bias was introduced simply by a spurious culture sample selection. Likewise, RD^{Rio} strains neither clustered in time over the 20 months studied nor segregated to a specific area in Rio. The latter findings support the notion that RD^{Rio} strains indeed are dispersed throughout the city and are responsible for a large proportion of the TB burden. Additional prospective population studies incorporating comprehensive reviews of patient data, as well as laboratory infection experiments, will be necessary to determine the relative virulence of RD^{Rio} strains compared to other *M. tuberculosis* genotypes. Nonetheless, the potential importance of the RD^{Rio} clade within the LAM family is underscored by the finding that the *M. tuberculosis* LAM clade contributes as much as 50% of the TB burden in South America and is prevalent in the Caribbean, in Mediterranean regions, and in parts of Africa (9).

In conclusion, this study provides a first impression of the *M. tuberculosis* population structure in Rio de Janeiro by a combined genotyping approach. This will serve as a platform for molecular epidemiological studies of the microevolution of endemic strains and the impact of persistent ancestral clones. The genetic definition of the *M. tuberculosis* RD^{Rio} genotype will aid in the identification of further isolates outside of Brazil and allow us to dissect the clinical, epidemiologic, and phenotypic traits unique to this breakthrough clade of strains.

ACKNOWLEDGMENTS

We are grateful to Dick van Soolingen, Michel Fabre, Gisele Betzler, Thiago T. Mafort, Pedro G. Macedo, Rodrigo L. L. Fernandes, André P. Shecaira, Fernanda S. César, Edson da Silva, S. Júnior e Cláudio, J. Valença, Atina R. Elias, Nalin Rastogi, Thierry Zozio, Scott Weisenberg, Alexandre Almeida, Christophe Sola, Afrânio L. Kritski, and Warren D. Johnson, Jr., for contributions to, and support of, this study.

Funding was provided by NIH grants R0-1 AI39606, R0-1 HL61960, R21 AI063147, and R21 AI063147 (to J.L.H.); NIH Fogarty International Center training grants (FICTG) (D43 TW00018); Innovative Approaches for TB Control in Brazil (ICOHRTA, U2R TW006885, J.R.L.E.S. and R. Chaisson [J.L.H.]) grants from the Coordenação de Aperfeicoamento de Pessoal de Nivel Superior (CAPES) (Ministry of

Education, Brazil), the Brazilian Ministry of Health (024/94 DST/AIDS), the Fundação Universitária José Bonifácio/FUJB, the Brazilian Research Council/CNPq, the Brazilian Research Council/World Bank Millennium Institute of Science, Programa de Apoio a Núcleos de Excelência (PRONEX), and PAPES IV/FIOCRUZ (to P.N.S. and J.R.L.E.S.); and a grant from the Laura Cook Hull Trust Fund (LCHTF) (Warren D. Johnson, Jr., principal investigator). R.C.H. was supported by the LCHTF, and L.C.O.L. was a FICTG and CAPES trainee.

REFERENCES

1. **Anh, D. D., M. W. Borgdorff, L. N. Van, N. T. Lan, T. van Gorkom, K. Kremer, and D. van Soolingen.** 2000. *Mycobacterium tuberculosis* Beijing genotype emerging in Vietnam. Emerg. Infect. Dis. **6:**302–305.
2. **Baptista, I. M., M. C. Oelemann, D. V. Opromolla, and P. N. Suffys.** 2002. Drug resistance and genotypes of strains of *Mycobacterium tuberculosis* isolated from human immunodeficiency virus-infected and non-infected tuberculosis patients in Bauru, Sao Paulo, Brazil. Mem. Inst. Oswaldo Cruz **97:**1147–1152.
3. **Barczak, A. K., P. Domenech, H. I. Boshoff, M. B. Reed, C. Manca, G. Kaplan, and C. E. Barry III.** 2005. In vivo phenotypic dominance in mouse mixed infections with *Mycobacterium tuberculosis* clinical isolates. J. Infect. Dis. **192:**600–606.
4. **Barnes, P. F., and M. D. Cave.** 2003. Molecular epidemiology of tuberculosis. N. Engl. J. Med. **349:**1149–1156.
5. **Beggs, M. L., K. D. Eisenach, and M. D. Cave.** 2000. Mapping of IS*6110* insertion sites in two epidemic strains of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **38:**2923–2928.
6. **Bifani, P. J., B. Mathema, N. E. Kurepina, and B. N. Kreiswirth.** 2002. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. Trends Microbiol. **10:**45–52.
7. **Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, M. C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole.** 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc. Natl. Acad. Sci. USA **99:**3684–3689.
8. **Brudey, K., I. Filliol, S. Ferdinand, V. Guernier, P. Duval, B. Maubert, C. Sola, and N. Rastogi.** 2006. Long-term population-based genotyping study of *Mycobacterium tuberculosis* complex isolates in the French departments of the Americas. J. Clin. Microbiol. **44:**183–191.
9. **Brudey, K., J. R. Driscoll, L. Rigouts, W. M. Prodinger, A. Gori, S. A. Al-Hajoj, C. Allix, L. Aristimuno, J. Arora, V. Baumanis, L. Binder, P. Cafrune, A. Cataldi, S. Cheong, R. Diel, C. Ellermeier, J. T. Evans, M. Fauville-Dufaux, S. Ferdinand, D. Garcia de Viedma, C. Garzelli, L. Gazzola, H. M. Gomes, M. C. Guttierez, P. M. Hawkey, P. D. van Helden, G. V. Kadival, B. N. Kreiswirth, K. Kremer, M. Kubin, S. P. Kulkarni, B. Liens, T. Lillebaek, M. L. Ho, C. Martin, C. Martin, I. Mokrousov, O. Narvskaia, Y. F. Ngeow, L. Naumann, S. Niemann, I. Parwati, Z. Rahim, V. Rasolofo-Razanamparany, T. Rasolonavalona, M. L. Rossetti, S. Rusch-Gerdes, A. Sajduda, S. Samper, I. G. Shemyakin, U. B. Singh, A. Somoskovi, R. A. Skuce, D. van Soolingen, E. M. Streicher, P. N. Suffys, E. Tortoli, T. Tracevska, V. Vincent, T. C. Victor, R. M. Warren, S. F. Yap, K. Zaman, F. Portaels, N. Rastogi, and C. Sola.** 2006. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol. **6:**23.
10. **Canetti, G., N. Rist, and J. Grosset.** 1963. Measurement of sensitivity of the tuberculous bacillus to antibacillary drugs by the method of proportions. Methodology, resistance criteria, results and interpretation. Rev. Tuberc. Pneumol. (Paris) **27:**217–272. (In French.)
11. **Caws, M., G. Thwaites, K. Stepniewska, N. T. N. Lan, N. T. H. Duyen, N. T. Phuong, M. N. T. Huyen, P. M. Duy, T. H. Loc, T. T. H. Chau, D. van Soolingen, K. Kremer, N. V. V. Chau, N. T. Chinh, and J. Farrar.** 2006. Beijing genotype of *Mycobacterium tuberculosis* is significantly associated with human immunodeficiency virus infection and multidrug resistance in cases of tuberculous meningitis. J. Clin. Microbiol. **44:**3934–3939.
12. **Clarridge, J. E., III.** 2004. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin. Microbiol. Rev. **17:**840–862.
13. **Cole, S. T.** 2002. Comparative and functional genomics of the *Mycobacterium tuberculosis* complex. Microbiology **148:**2919–2928.
14. **Corbett, E. L., C. J. Watt, N. Walker, D. Maher, B. G. Williams, M. C. Raviglione, and C. Dye.** 2003. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. Arch. Intern. Med. **163:**1009–1021.
15. **Cowan, L. S., L. Mosher, L. Diem, J. P. Massey, and J. T. Crawford.** 2002. Variable-number tandem repeat typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110* by using mycobacterial interspersed repetitive units. J. Clin. Microbiol. **40:**1592–1602.
16. **Dale, J. W., D. Brittain, A. A. Cataldi, D. Cousins, J. T. Crawford, J. Driscoll, H. Heersma, T. Lillebaek, T. Quitugua, N. Rastogi, R. A. Skuce, C. Sola, D. Van Soolingen, and V. Vincent.** 2001. Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standardized nomenclature. Int. J. Tuberc. Lung Dis. **5:**216–219.
17. **Drobniewski, F., Y. Balabanova, V. Nikolayevsky, M. Ruddy, S. Kuznetzov, S. Zakharova, A. Melentyev, and I. Fedorin.** 2005. Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia. JAMA **293:**2726–2731.
18. **Fang, Z., C. Doig, D. T. Kenna, N. Smittipat, P. Palittapongarnpim, B. Watt, and K. J. Forbes.** 1999. IS*6110*-mediated deletions of wild-type chromosomes of *Mycobacterium tuberculosis*. J. Bacteriol. **181:**1014–1020.
19. **Ferdinand, S., G. Valetudie, C. Sola, and N. Rastogi.** 2004. Data mining of *Mycobacterium tuberculosis* complex genotyping results using mycobacterial interspersed repetitive units validates the clonal structure of spoligotyping-defined families. Res. Microbiol. **155:**647–654.
20. **Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. A. Dang, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F. Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, E. Kassa-Kelembho, M. L. Ho, A. Makristathis, C. Mammina, G. Martin, P. Mostrom, I. Mokrousov, V. Narbonne, O. Narvskaya, A. Nastasi, S. N. Niobe-Eyangoh, J. W. Pape, V. Rasolofo-Razanamparany, M. Ridell, M. L. Rossetti, F. Stauffer, P. N. Suffys, H. Takiff, J. Texier-Maugein, V. Vincent, J. H. de Waard, C. Sola, and N. Rastogi.** 2003. Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. J. Clin. Microbiol. **41:**1963–1970.
21. **Fleischmann, R. D., D. Alland, J. A. Eisen, L. Carpenter, O. White, J. Peterson, R. DeBoy, R. Dodson, M. Gwinn, D. Haft, E. Hickey, J. F. Kolonay, W. C. Nelson, L. A. Umayam, M. Ermolaeva, S. L. Salzberg, A. Delcher, T. Utterback, J. Weidman, H. Khouri, J. Gill, A. Mikula, W. Bishai, W. R. Jacobs, Jr., J. C. Venter, and C. M. Fraser.** 2002. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. J. Bacteriol. **184:**5479–5490.
22. **Frothingham, R., P. L. Strickland, G. Bretzel, S. Ramaswamy, J. M. Musser, and D. L. Williams.** 1999. Phenotypic and genotypic characterization of *Mycobacterium africanum* isolates from West Africa. J. Clin. Microbiol. **37:**1921–1926.
23. **Gagneux, S., K. DeRiemer, T. Van, M. Kato-Maeda, B. C. Jong, S. Narayanan, M. Nicol, S. Niemann, K. Kremer, M. C. Gutierrez, M. Hilty, P. C. Hopewell, and P. M. Small.** 2006. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. USA **103:**2869–2873.
24. **Garnier, T., K. Eiglmeier, J. C. Camus, N. Medina, H. Mansoor, M. Pryor, S. Duthoy, S. Grondin, C. Lacroix, C. Monsempe, S. Simon, B. Harris, R. Atkin, J. Doggett, R. Mayes, L. Keating, P. R. Wheeler, J. Parkhill, B. G. Barrell, S. T. Cole, S. V. Gordon, and R. G. Hewinson.** 2003. The complete genome sequence of *Mycobacterium bovis*. Proc. Natl. Acad. Sci. USA **100:**7877–7882.
25. **Glynn, J. R., A. C. Crampin, M. D. Yates, H. Traore, F. D. Mwaungulu, B. M. Ngwira, R. Ndlovu, F. Drobniewski, and P. E. Fine.** 2005. The importance of recent infection with *Mycobacterium tuberculosis* in an area with high HIV prevalence: a long-term molecular epidemiological study in Northern Malawi. J. Infect. Dis. **192:**480–487.
26. **Goh, K. S., N. Rastogi, M. Berchel, R. C. Huard, and C. Sola.** 2005. Molecular evolutionary history of tubercle bacilli assessed by study of the polymorphic nucleotide within the nitrate reductase (*narGHJI*) operon promoter. J. Clin. Microbiol. **43:**4010–4014.
27. **Gutacker, M. M., B. Mathema, H. Soini, E. Shashkina, B. N. Kreiswirth, E. A. Graviss, and J. M. Musser.** 2006. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. J. Infect. Dis. **193:**121–128.
28. **Heersma, H. F., K. Kremer, and J. D. van Embden.** 1998. Computer analysis of IS*6110* RFLP patterns of *Mycobacterium tuberculosis*. Methods Mol. Biol. **101:**395–422.
29. **Ho, T. B., B. D. Robertson, G. M. Taylor, R. J. Shaw, and D. B. Young.** 2000. Comparison of *Mycobacterium tuberculosis* genomes reveals frequent deletions in a 20 kb variable region in clinical isolates. Yeast **17:**272–282.
30. **Huard, R. C., L. C. Lazzarini, W. R. Butler, D. van Soolingen, and J. L. Ho.** 2003. A PCR-based method to differentiate the subspecies of the *Mycobacterium tuberculosis* complex on the basis of genomic deletions. J. Clin. Microbiol. **41:**1637–1650.
31. **Huard, R. C., M. Fabre, P. de Haas, L. C. Lazzarini, D. van Soolingen, D. Cousins, and J. L. Ho.** 2006. Novel genetic polymorphisms that further delineate the phylogeny of the *Mycobacterium tuberculosis* complex. J. Bacteriol. **188:**4271–4287.
32. **Ignatova, A., S. Dubiley, V. Stepanshina, and I. Shemyakin.** 2006. Predominance of multi-drug-resistant LAM and Beijing family strains among *Mycobacterium tuberculosis* isolates recovered from prison inmates in Tula Region, Russia. J. Med. Microbiol. **55:**1413–1418.
33. **Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal, and J. van Embden.** 1997. Simultaneous detection and strain differentiation of *Myco-*

bacterium tuberculosis for diagnosis and epidemiology. J. Clin. Microbiol. **35:**907–914.
34. **Karboul, A., N. C. Gey van Pittius, A. Namouchi, V. Vincent, C. Sola, N. Rastogi, P. Suffys, M. Fabre, A. Cataldi, R. C. Huard, N. Kurepina, B. N. Kreiswirth, J. L. Ho, M. C. Gutierrez, and H. Mardassi.** 2006. Insights into the evolutionary history of tubercle bacilli as disclosed by genetic rearrangements within a PE_PGRS duplicated gene pair. BMC Evol. Biol. **6:**107.
35. **Kong, Y., M. D. Cave, L. Zhang, B. Foxman, C. F. Marrs, J. H. Bates, and Z. H. Yang.** 2006. Population-based study of deletions in five different genomic regions of *Mycobacterium tuberculosis* and possible clinical relevance of the deletions. J. Clin. Microbiol. **44:**3940–3946.
36. **Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. M. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. A. van Embden.** 1999. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J. Clin. Microbiol. **37:**2607–2618.
37. **Kurepina, N. E., S. Sreevatsan, B. B. Plikaytis, P. J. Bifani, N. D. Connell, R. J. Donnelly, D. van Soolingen, J. M. Musser, and B. N. Kreiswirth.** 1998. Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS*6110* elements in *Mycobacterium tuberculosis*: non-random integration in the *dnaA-dnaN* region. Tuber. Lung Dis. **79:**31–42.
38. **López, B., D. Aguilar, H. Orozco, M. Burger, C. Espitia, V. Ritacco, L. Barrera, K. Kremer, R. Hernandez-Pando, K. Huygen, and D. van Soolingen.** 2003. A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes. Clin. Exp. Immunol. **133:**30–37.
39. **Macaraig, M., T. Agerton, C. R. Driver, S. S. Munsiff, J. Abdelwahab, J. Park, B. Kreiswirth, J. Driscoll, and B. Zhao.** 2006. Strain-specific differences in two large *Mycobacterium tuberculosis* genotype clusters in isolates collected from homeless patients in New York City from 2001 to 2004. J. Clin. Microbiol. **44:**2890–2896.
40. **Marmiesse, M., P. Brodin, C. Buchrieser, C. Gutierrez, N. Simoes, V. Vincent, P. Glaser, S. T. Cole, and R. Brosch.** 2004. Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. Microbiology **150:**483–496.
41. **McNabb, A., D. Eisler, K. Adie, M. Amos, M. Rodrigues, G. Stephens, W. A. Black, and J. Isaac-Renton.** 2004. Assessment of partial sequencing of the 65-kilodalton heat shock protein gene (*hsp65*) for routine identification of *Mycobacterium* species isolated from clinical sources. J. Clin. Microbiol. **42:**3000–3011.
42. **Mostowy, S., D. Cousins, J. Brinkman, A. Aranaz, and M. A. Behr.** 2002. Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. J. Infect. Dis. **186:**74–80.
43. **Newton, S. M., R. J. Smith, K. A. Wilkinson, M. P. Nicol, N. J. Garton, K. J. Staples, G. R. Stewart, J. R. Wain, A. R. Martineau, S. Fandrich, T. Smallie, B. Foxwell, A. Al-Obaidi, J. Shafi, K. Rajakumar, B. Kampmann, P. W. Andrew, L. Ziegler-Heitbrock, M. R. Barer, and R. J. Wilkinson.** 2006. A deletion defining a common Asian lineage of *Mycobacterium tuberculosis* associates with immune subversion. Proc. Natl. Acad. Sci. USA **103:**15594–15598.
44. **Niobe-Eyangoh, S. N., C. Kuaban, P. Sorlin, J. Thonnon, V. Vincent, and M. C. Gutierrez.** 2004. Molecular characteristics of strains of the Cameroon family, the major group of *Mycobacterium tuberculosis* in a country with a high prevalence of tuberculosis. J. Clin. Microbiol. **42:**5029–5035.
45. **Okie, S.** 2006. Fighting HIV—lessons from Brazil. N. Engl. J. Med. **354:**1977–1981.
46. **Reed, M. B., P. Domenech, C. Manca, H. Su, A. K. Barczak, B. N. Kreiswirth, G. Kaplan, and C. E. Barry III.** 2004. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature **431:**84–87.
47. **Salamon, H., M. Kato-Maeda, P. M. Small, J. Drenkow, and T. R. Gingeras.** 2000. Detection of deleted genomic DNA using a semiautomated computational analysis of GeneChip data. Genome Res. **10:**2044–2054.
48. **Scorpio, A., and Y. Zhang.** 1996. Mutations in *pncA*, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. Nat. Med. **2:**662–667.
49. **Singh, K. K., Y. Dong, S. A. Patibandla, D. N. McMurray, V. K. Arora, and S. Laal.** 2005. Immunogenicity of the *Mycobacterium tuberculosis* PPE55 (Rv3347c) protein during incipient and clinical tuberculosis. Infect. Immun. **73:**5004–5014.
50. **Sola, C., I. Filliol, E. Legrand, I. Mokrousov, and N. Rastogi.** 2001. *Mycobacterium tuberculosis* phylogeny reconstruction based on combined numerical analysis with IS*1081*, IS*6110*, VNTR, and DR-based spoligotyping suggests the existence of two new phylogeographical clades. J. Mol. Evol. **53:**680–689.
51. **Sola, C., I. Filliol, E. Legrand, S. Lesjean, C. Locht, P. Supply, and N. Rastogi.** 2003. Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. Infect. Genet. Evol. **3:**125–133.
52. **Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth, T. S. Whittam, and J. M. Musser.** 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. USA **94:**9869–9874.
53. **Suffys, P. N., M. E. Ivens de Araujo, M. L. Rossetti, A. Zahab, E. W. Barroso, A. M. Barreto, E. Campos, D. van Soolingen, K. Kremer, H. Heersma, and W. M. Degrave.** 2000. Usefulness of IS*6110*-restriction fragment length polymorphism typing of Brazilian strains of *Mycobacterium tuberculosis* and comparison with an international fingerprint database. Res. Microbiol. **151:**343–351.
54. **Suffys, P. N., M. E. Ivens de Araújo, and W. M. Degrave.** 1997. The changing face of the epidemiology of tuberculosis due to molecular strain typing—a review. Mem. Inst. Oswaldo Cruz **92:**297–316.
55. **Supply, P., C. Allix, S. Lesjean, M. Cardoso-Oelemann, S. Rusch-Gerdes, E. Willery, E. Savine, P. de Haas, H. van Deutekom, S. Roring, P. Bifani, N. Kurepina, B. Kreiswirth, C. Sola, N. Rastogi, V. Vatin, M. C. Gutierrez, M. Fauville, S. Niemann, R. Skuce, K. Kremer, C. Locht, and D. van Soolingen.** 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **44:**4498–4510.
56. **Supply, P., S. Lesjean, E. Savine, K. Kremer, D. van Soolingen, and C. Locht.** 2001. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. J. Clin. Microbiol. **39:**3563–3571.
57. **Tuberculosis Control Program of the City of Rio de Janeiro.** 2005. Boletim informativo do programa de controle de TB do municipio do Rio de Janeiro. http://www.saude.rio.rj.gov.br/media/boletim_epidemiologicoTB_2004.pdf.
58. **van Embden, J. D., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. M. Shinnick, and P. M. Small.** 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. **31:**406–409.
59. **van Soolingen, D., L. Qian, P. E. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. van Embden.** 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J. Clin. Microbiol. **33:**3234–3238.
60. **Talaat, A. M., R. Lyons, S. T. Howard, and S. A. Johnston.** 2004. The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. Proc. Natl. Acad. Sci. USA **101:**4602–4607.
61. **Voskuil, M. I., D. Schnappinger, R. Rutherford, Y. Liu, and G. K. Schoolnik.** 2004. Regulation of the *Mycobacterium tuberculosis* PE/PPE genes. Tuberculosis **84:**256–262.
62. **Warren, R. M., E. M. Streicher, S. L. Sampson, G. D. van der Spuy, M. Richardson, D. Nguyen, M. A. Behr, T. C. Victor, and P. D. van Helden.** 2002. Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: implications for interpretation of spoligotyping data. J. Clin. Microbiol. **40:**4457–4465.
63. **Warren, R. M., T. C. Victor, E. M. Streicher, M. Richardson, N. Beyers, N. C. Gey van Pittius, and P. D. van Helden.** 2004. Patients with active tuberculosis often have different strains in the same sputum specimen. Am. J. Respir. Crit. Care Med. **169:**610–614.
64. **World Health Organization.** 2000. Tuberculosis fact sheet no. 104. World Health Organization, Geneva, Switzerland. http://www.who.int/mediacentre/factsheets/fs104/en/index.html.
65. **World Health Organization.** 2007. WHO 2007 report. Global tuberculosis control. Country profile: Brazil. World Health Organization, Geneva, Switzerland. http://www.who.int/GlobalAtlas/predefinedReports/TB/PDF_Files/bra.pdf.
66. **World Health Organization.** 2007. Global health atlas. World Health Organization, Geneva, Switzerland. http://www.who.int/globalatlas/.